

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Patent Application of)	
)	
Scott Shepard et al.)	Group Art Unit: 2626
)	
Application No.: 10/610,684)	Examiner: Matthew H. Baker
)	
Filed: July 2, 2003)	
)	
For: SYSTEMS AND METHODS FOR)	
AIDING HUMAN TRANSLATION)	

APPEAL BRIEF

Mail Stop: APPEAL BRIEF - PATENT
Commissioner for Patents
P.O. Box 1450
Alexandria, Virginia 22313-1450

Sir:

This Appeal Brief is filed under Rule 41.37, appealing the final decision of the Examiner dated February 3, 2009 and is further responsive to the Advisory Action dated May 20, 2009. A Notice of Appeal was filed on July 17, 2009 along with an appropriate petition for extension of time for response. Each of the topics required by Rule 41.37 is presented herewith and labeled in accordance therewith. Only one copy of this Appeal Brief is required, in accordance with MPEP §1205.02.

TABLE OF CONTENTS

REAL PARTY IN INTEREST.....	PAGE 3
RELATED APPEALS AND INTERFERENCES.....	PAGE 4
STATUS OF CLAIMS.....	PAGE 5
STATUS OF AMENDMENTS.....	PAGE 6
SUMMARY OF CLAIMED SUBJECT MATTER.....	PAGES 7-11
GROUND OF REJECTION TO BE REVIEWED ON APPEAL.....	PAGE 12
ARGUMENTS.....	PAGES 13-33
FOSTER DISCLOSURE.....	PAGE 13-14
SCHULZ DISCLOSURE.....	PAGE 14
SHIOTANI DISCLOSURE.....	PAGE 14-15
OVERVIEW	PAGES 15-17
CLAIM 1 ARGUMENTS.....	PAGES 17-28
CLAIM 47 ARGUMENTS.....	PAGES 28-34
CONCLUSION.....	PAGE 35
CLAIMS APPENDIX.....	PAGES 36-48
EVIDENCE APPENDIX.....	PAGE 49
RELATED PROCEEDINGS APPENDIX.....	PAGE 50

I. REAL PARTY IN INTEREST

The real party in interest of the present application, solely for purposes of identifying and avoiding potential conflicts of interest by board members due to working in matters in which the member has a financial interest, is Verizon Communications Inc. and its subsidiary companies, which currently include Verizon Business Global, LLC (formerly MCI, LLC) and Cellco Partnership (doing business as Verizon Wireless, and which includes as a minority partner affiliates of Vodafone Group Plc). Verizon Communications Inc. or one of its subsidiary companies is an assignee of record of the present application.

II. RELATED APPEALS AND INTERFERENCES

There are no appeals or interferences related to the present application of which the Appellants are aware.

III. STATUS OF CLAIMS

Claims 1-47 are currently pending in the application and all stand finally rejected.

Claims 1-47 are identified as claims that are being appealed.

IV: STATUS OF AMENDMENTS

Subsequent to the final Office Action of February 3, 2009, (hereinafter "final Office Action"), Appellants filed an after-final Reply under 37 C.F.R. §1.116 but did not amend the claims. The last amendment in this application was filed on September 16, 2008 responsive to a non-final office action dated June 27, 2008. Accordingly, there are no outstanding amendments in this application.

V. SUMMARY OF CLAIMED SUBJECT MATTER

The following summary of the presently claimed subject matter indicates that certain portions of the specification (including the drawings) provide examples of embodiments of elements of the claimed subject matter. It is to be understood that other portions of the specification not cited herein may also provide examples of embodiments of elements of the claimed subject matter. It is also to be understood that the indicated examples are merely examples, and the scope of the claimed subject matter includes alternative embodiments and equivalents thereof. References herein to the specification are thus intended to be exemplary and not limiting.

In overview of the claimed subject matter, Appellants teach a human-language translation system in which a human-being translator does the translating from a first language (e.g., Spanish) to a second language (e.g., English). The human translator listens to an audio message in the first language while reading a text-transcript of that same audio message in that first language on a portion of a split screen display. The audio message and its corresponding text-transcript are to be translated by that person into the second language, in this example from Spanish to English. Each word in the text transcript of that first language is highlighted on the display screen in synchrony with the utterance of that word as it is spoken in the corresponding audio message. This serves as a translation-aid to the human translator as he, or she, types the translated message in the second language on another portion of the split screen. (See, e.g., Fig. 10) There is no machine-translating involved in Appellants' claimed subject matter - only machine-transcribing (transcribing from audio to text in the same language). With this overview

in mind, consider the claimed subject matter in detail. Appellants hereby map all independent claims to the drawings and Specification.

Independent claim 1 recites a method for facilitating translation of an audio signal that includes speech to another language, (e.g., at least Specification ¶'s [0008], [0009] and [0010] and Fig. 10) comprising:

retrieving a textual representation of the audio signal; (e.g., at least Specification ¶ [0044] and Fig. 4)

presenting the textual representation to a user; (e.g., at least Specification ¶'s [0045], [0046] and [0047] and Figs. 4-5)

receiving selection of a segment of the textual representation for translation; (e.g., at least Specification ¶ [0052] and Fig. 4)

obtaining a portion of the audio signal corresponding to the segment of the textual representation; (e.g., at least Specification ¶'s [0053] and [0059] and Figs. 4 and 8)

providing the segment of the textual representation and the portion of the audio signal to the user; (e.g., at least Specification ¶'s [0054] and [0065] and Figs. 4 and 8) and

receiving translation actually made by the user of the portion of the audio signal (e.g., at least Specification ¶'s [0005], [0008], [0009], [0066], [0070]-[0072] and Figs. 8-10).

Independent claim 20 recites a system for facilitating translation of speech between languages, (e.g., at least Specification ¶'s [0008], [0009] and [0010] and Fig. 10) comprising:

means for obtaining a textual representation of the speech in a first language (e.g., at least Specification ¶ [0044] and Fig. 4);

means for presenting the textual representation to a user (e.g., at least Specification ¶'s [0045], [0046] and [0047] and Figs. 4-5);

means for receiving selection of a portion of the textual representation for translation (e.g., at least Specification ¶ [0052] and Fig. 4);

means for retrieving an audio signal in the first language that corresponds to the portion of the textual representation (e.g., at least Specification ¶'s [0053] and [0059] and Figs. 4 and 8);

means for providing the portion of the textual representation and the audio signal to the user (e.g., at least Specification ¶'s [0054] and [0065] and Figs. 4 and 8); and

means for receiving translation actually made by the user of the audio signal into a second language (e.g., at least Specification ¶'s [0005], [0008], [0009], [0066], [0070]-[0072] and Figs. 8-10).

Independent claim 21 recites a translation system (e.g., at least Specification ¶'s [0008] - [0010] and [0024] - [0025] and Figs. 1-3 and 10), comprising:

a memory configured to store instructions (e.g., at least Specification ¶ [0028] and Fig. 2); and

a processor configured to execute the instructions in memory (e.g., at least Specification ¶ [0028] and Fig. 2) to:

obtain a transcription of an audio signal that includes speech (e.g., at least

Specification ¶ [0044] and Fig. 4),

present the transcription to a user (e.g., at least Specification ¶'s [0045], [0046] and [0047] and Figs. 4-5),

receive selection of a portion of the transcription for translation (e.g., at least Specification ¶ [0052] and Fig. 4),

retrieve a portion of the audio signal corresponding to the portion of the transcription (e.g., at least Specification ¶'s [0053] and [0059] and Figs. 4 and 8),

provide the portion of the transcription and the portion of the audio signal to the user (e.g., at least Specification ¶'s [0054] and [0065] and Figs. 4 and 8), and

receive from the user a translation actually made by the user of the portion of the audio signal (e.g., at least Specification ¶'s [0005], [0008], [0009], [0066], [0070]-[0072] and Figs. 8-10).

Independent claim 40 recites a graphical user interface (e.g., at least Specification ¶'s [0061], [0062] and [0066] and Figs. 9-10), comprising:

a transcription section that includes a transcription of non-text information in a first language (e.g., at least Specification ¶ [0062] and Figs. 9-10);

a translation section that receives a translation actually made by the user of the non-text information into a second language (e.g., at least Specification ¶'s [0005], [0008], [0009], [0062], [0066], [0070]-[0072] and Figs. 8-10); and

a play button (e.g., at least Specification ¶ [0062] and Figs. 9-10) that, when selected,

causes:

retrieval of the non-text information to be initiated (e.g., at least Specification ¶'s [0063] - [0064] and Figs. 8-10),

playing of the non-text information (e.g., at least Specification ¶ [0064] and Figs. 8-10), and

the playing of the non-text information to be visually synchronized with the transcription in the transcription section (e.g., at least Specification ¶ [0065] and Figs. 8-10).

Independent claim 47 recites a method (e.g., at least Figs. 4 and 8), comprising:

a user listening to an audio playback of information in a first language while viewing a textual transcription of said information in said first language on a transcription section of a graphical user interface (GUI), said textual transcription being synchronized with said audio playback (e.g., at least Specification ¶'s [0065] - [0066] and Fig. 10); and

said user actually translating said audio playback of said information thereby obtaining a translation in a second language, said user using a different section of said GUI to display said translation while making said translation (e.g., at least Specification ¶'s [0005], [0008], [0009], [0065], [0066], [0070]-[0072] and Figs. 8-10),

whereby the synchronizing of said audio playback with said textual transcription aids said user in making said translation (e.g., at least Specification ¶'s [0070] - [0071]).

VL. GROUNDS OF REJECTION TO BE REVIEWED ON APPEAL

In the final Office Action, where the following three rejections were made, rejections one and three are the only grounds of rejection to be reviewed on appeal:¹

Rejection Number One (to be reviewed on appeal):

Claims 1-11, 13-31, 33-38, 40 and 44-45 are rejected under 35 U.S.C. § 103 (a) as being un-patentable over Foster (“Target-Text Mediated Interactive Machine Translation” Machine Translation, 1997 and hereinafter referred to as “Foster”) in view of U.S. Patent No. 6,360,237 to Schulz et al. (hereinafter, “Schulz”).

Rejection Number Two: (not reviewed on appeal)

Claims 41 and 46 are rejected under 35 U.S.C. § 103(a) as being un-patentable over Foster in view of Schulz, and further in view of U.S. Patent No. 6,820,055 to Saindon et al. (hereinafter, “Saindon”).

Rejection Number Three: (to be reviewed on appeal)

Claims 12, 19, 32, 39, 42, 43 and 47 are rejected under 35 U.S.C. § 103(a) as being un-patentable over Foster in view of Schulz as applied to claims 1, 21 and 40 and further in view of U.S. Patent No. 4,814,988 to Shiotani et al. (hereinafter “Shiotani”).

¹ As noted in the Argument section which follows, all independent claims but for claim 47 shall stand or fall with claim 1. Furthermore, Appellants’ dependent claims shall stand or fall with their independent claims. A separate argument is presented for independent claim 47. Thus, the second ground of rejection is moot.

VII: ARGUMENT

The independent claims on appeal are claims 1, 20, 21, 40 and 47. Appellants shall let the dependent claims stand or fall with their respective independent claims and let independent claims 20, 21, and 40 stand or fall with claim 1. An additional argument is presented for independent claim 47. Therefore, the only claims for which arguments are being presented below are claims 1 and 47. The second ground of rejection is moot.

SUMMARY OF THE FOSTER DISCLOSURE:

Foster discloses target-text mediated interactive machine translation. (Title) Thus, it teaches translation of text only (no audio) by a human translator in combination with a machine. Foster teaches that a human translator can start the translation of a word appearing in a source human language by entering a keystroke and by entering follow-on, consecutive keystrokes. (pg 179, section 3. "Word Completion") At some point after the first keystroke, the machine can suggest a completed word in the target language, presumably corresponding, or close-in-meaning, to the word in the source language. (pg 179, section 3. "Word Completion") The machine's translation-offering is based on the thought process of the human translator which is reflected in his/her keystroke(s). In turn, the human translator considers the machine's translation offering and either accepts the word or, if inappropriate in his/her judgment, makes the next keystroke thereby providing the next letter in that word.

Therefore, the next keystroke by the human translator is, of necessity, based on (1) his/her understanding of the meaning being expressed in human source language and (2) his/her

view of the translated word offered by the machine. An inappropriate offering from the machine in his/her view necessarily serves to at least reinforce the word choice then being contemplated by the human translator, if not to actually guide him/her more swiftly toward that word choice. In this manner, the machine may offer a different word candidate based on each succeeding keystroke made by the human translator and thereby contribute to the translation process.

The human translator shall probably not complete each word solely by his/her keystrokes. Indeed, the Foster disclosure (pg 192) estimates that the machine could reduce the number of keystrokes by approximately 70% and, therefore, is reasonably likely to propose an acceptable word prior to the human translator finishing the translation of a word by himself/herself. Regardless, the machine contributes to the translation process of each and every word.

SUMMARY OF THE SCHULZ DISCLOSURE:

Schulz relates to a method and system for performing text edits during audio recording playback for transcription. (title and col. 1, lines 7-9) Schulz does not disclose, or relate to, translation from one language to a different language but is limited to transcribing from audio to text within the same language. Schulz discloses a method for editing (correcting) written text in a particular language in a text editor which automatically aligns a cursor in the written text on a screen with a particular spoken word in that same language during playback of an audio recording. (col. 2, lines 48-51) The text editor is a software application. (col. 4, line 24) A human transcriptionist or user may edit the text using special edit function keys. (col. 5, lines 44-

45) Thus, Schulz may show synchronization between audio and corresponding text, in the same human source language, without any translation involved. There is no translation taught or suggested in Schulz.

SUMMARY OF THE SHIOTANI DISCLOSURE:

Shiotani relates to a machine translation system for translating all or a selected portion of an input sentence. (title) The Shiotani input is derived from non-audio sources such as an optical character reader (OCR). (col. 2, lines 11-14) Shiotani, therefore, does not teach translation of language presented in an audio format. Shiotani presents a block diagram of essential parts of its machine translation system in its Fig. 1. (col. 1, lines 66-67). Translating part 7, shown in Shiotani's Fig. 1, is the computer mechanism that does the actual translating; it translates the content in original buffer 6 by operation of a dictionary look-up/morpheme analyzing function, a syntax analyzing function, a transforming function and a generating function. (col. 2, lines 30-34) A translation buffer 10 is provided for storing the result of the translation. (col. 2, lines 38-39) In addition, a correcting means 11 is used by the human operator to correct the translation result that is displayed on a terminal screen. (col. 2, lines 38-41) Accordingly, Shiotani teaches machine translation of non-audio human language and human correction of that translation.

I. OVERVIEW:

The Advisory Action, pg 2, paragraph 5, argues that "Applicant should direct attention to Section 4-1 [Foster] where it is made clear that 'the computer assists the human, rather than vice

versa,' in other words the human is doing the translating while the machine offers [translation] suggestions." It is agreed that the computer provides assistance to the human, but this is translation assistance. In addition, section 4-2 says: "the human translator issues directives in the form of characters, words, or possibly more abstract properties, and the computer reacts to each with a revised proposal for all or part of the target text." Thus, the machine proposes translated words in reaction to each "directive" and, not surprisingly, is also translating (after all, the title of Foster includes the phrase "Machine Translation"). This is the main point of Appellants' argument - that the human translator in Foster does not do it alone - ever.

Even if none of the machine-translated word offerings were, *arguendo*, ever accepted, (and such performance is not taught as Foster discloses an estimated 70% success rate for the machine) the machine would, nevertheless, loyally persevere on every keystroke, without fail, to offer a translated word that may work for the human translator. This is a true cooperative effort between man and machine - a partnership.

To gain a true perspective, momentarily step back from this man/machine partnership. Hypothetically, if a reference had taught, instead of a machine, a second human translator who offered a translated word for every keystroke made by a first human translator, for acceptance by the first human translator, would that cooperative process be viewed as a translation actually made by the first human translator? No. The second translator's input is a true word-translation input and always reinforces the first translator's input. This is true even when not accepted by the first human translator because to the extent the second input shows less than an optimum translation-direction, it thereby necessarily channels the translation effort in the optimum

translation-direction by process of word-elimination. Thus, two people are translating and one of them makes the final decision on a per word basis while an ever-present translation input is made by the second human translator, even when a translated word offering is not accepted.

The only difference between the hypothetical above and the disclosure of Foster is a second human partner vs. a machine partner. Therefore, it is clear that Foster is not a reference which shows a “translation actually made by the user of the portion of the audio signal” as recited in claim 1 because the Foster machine always contributes to the translation made by both.²

I. CLAIM 1 IS ALLOWABLE BECAUSE FOSTER AND SCHULZ DO NOT DISCLOSE OR SUGGEST ALL CLAIM ELEMENTS

Claim 1 recites, *inter-alia*: “receiving translation actually made by the user of the portion of the audio signal”. Foster and Schulz taken individually or in any reasonable combination do not disclose or suggest this limitation for the following reasons:

Foster:

Principal reference Foster relates to target-text mediated interactive machine translation. (title) It relates to translation of text. It does not disclose or suggest translation of audio. Although Foster does involve a human translator, as Appellants shall explain below, that particular human involvement is not sufficient to enable Foster to be read on Appellants’ claim 1.

² Extrapolating, any translation-related document describing technology that assists a human translator automatically has no value as an effective reference against “translation actually made by the [human] user of the portion of the audio signal” as recited in claim 1.

The Office Action's reliance on Foster is limited to page 179, section 3, first paragraph. It uses this cited portion of Foster to allegedly read on Appellants' claim 1 and repeatedly uses only this cited portion of Foster to allegedly read on Appellants' other independent claims (claims 20 on pg 6, 21 on pg 8, 40 on pg 19, and 47 on pg 23). The cited section says:

"Our word-completion system works as follows: a translator selects some portion of the source text, nominally a sentence, and begins typing its translation. After each character is entered, the system displays a proposed completion for the current word, which the translator may either accept using a special command or reject by continuing to type. We chose this interface for our initial prototype because it is simple and because it allows performance to be measured easily by counting the proportion of characters or keystrokes saved in a test corpus; these are statistics that seem likely to correlate well with actual savings in human effort."

(Foster, pg. 179, section 3, paragraph 1, emphasis added) This section is saying that a human translator can select a sentence of source-language text and can begin typing other text in a target language based on the source text that he/she is reading. The human translator initiates the process by beginning to type letters which, if carried to completion, would spell a target-language-equivalent of the first word in the selected sentence. If the human translator agrees with the machine's proposed completion of that word in the displayed-text target language, he/she can accept it; if not, he/she continues to type (letter-by-letter) that first word in the target language, and the machine offers a new proposed completion after each typed letter unless and until the machine "gets it right" in the opinion of the human operator whereupon the human operator accepts the machine's translation input.

Therefore, Foster teaches a human-machine translation partnership. The Board is respectfully referred to Foster, page 177, paragraph at bottom of page:

"TTM [Target-Text Mediation] can in principle accommodate a wide range of MT proficiencies. Simple systems would be of benefit mainly in speeding the

transcription of the translator's work; more capable ones would add to this the occasional ability to suggest solutions that may otherwise have eluded (at least temporarily) THE HUMAN PARTNER."

(Foster, page 177, bottom, emphasis added) Therefore, Foster itself correctly views, and teaches, that its combination of machine translation and human translation is a partnership activity. To ignore this portion of the Foster disclosure does not comport with MPEP 2141.02(VI) which requires that a prior art reference must be considered in its entirety including portions that would lead away from the claimed subject matter.

The human translator receives the machine's input and decides whether or not to accept it. The machine is a partner in the translation effort of every word in the portion to be translated. The human translator is the judge of the accuracy or acceptability of the machine's translation and, if the translator agrees with the machine's input, he/she can then accept the machine's translation to enable the word to be completely translated possibly more quickly than otherwise. But, even if the machine's offering is rejected by the human translator as being other than the optimum translated word, he/she still uses that rejected machine input in a positive way to mentally rule-out a translation direction suggested by that rejected word as supplied by the machine. This helps the human translator in his/her mental process to more quickly select the optimum translated word in the target language, which the human translator is seeking.

In other words, even when the Foster machine attempts to finalize a word with a less than optimum, or plainly wrong, choice in the view of the human translator, the machine is still working together with the human translator on translating each and every word. It is clear that Foster teaches a human-machine translating partnership, or joint effort, and therefore cannot read

on “receiving translation actually made by the user of the portion of the audio signal.”

(emphasis added) Quite differently, in Foster, the translation process is actually made by the human-machine partnership all the time, whether or not the machine finishes a translated word correctly.

Translation is not only the end result, it is also the process by which the end result is reached. The relevant dictionary definition of “translation” is: “1: an act, process, or instance of translating: as a : a rendering from one language to into another; also: the product of such a rendering” and of “actually” is: “in act or in fact: REALLY; in point of fact: in truth - used to suggest something unexpected.”³ Clearly, based on ordinary usage as expressed in these dictionary definitions, the Foster-Schulz combination does not teach “receiving translation actually made by the user of the portion of the audio signal” as recited in claim 1 at least because “in fact” the translation was not “really” made by the human translator, but was made by the translator-machine partnership.

The entire reason for the existence of Foster is to have a machine help in the translation all the time, whether or not the machine supplies words acceptable to the human translator. Even if a word is rejected by the human translator, the machine-translation effort, nevertheless, was made. Indeed, there is a possibility that the machine could, in a particular word instance, be even more correct than the human translator who rejects a perfect translation offering which he/she didn't appreciate at that moment. Thus, Appellants see Foster as disclosing and suggesting a machine-human translation partnership where it is not possible for Foster to read

³ Merriam Webster's Collegiate Dictionary, Tenth Edition

on: "receiving translation actually made by the user..." as recited in claim 1. Rather, Foster teaches translation actually made by a user-machine partnership. This does not read on Appellants' claim limitation. Appellants believe this interpretation to be the proper interpretation and view of Foster relative to Appellants' claim limitation.

Thus, Appellants' position is that Foster does not read on Appellants' claim limitation: "receiving translation actually made by the user of the portion of the audio signal" as recited in claim 1, even if the human translator never accepts any translation offerings from the machine. However, Foster also discloses that it is not likely that that would happen. If Foster performs as it is estimated, then it can assist a human translator by completing the translation started by reducing the number of keystrokes needed to type target text words by approximately 70%. (Foster page 192).

This aspect of Foster's disclosure must be considered and weighed by the Examiner. As noted in MPEP 2141.02(VI), a prior art reference must be considered in its entirety, i.e., as a whole, including portions that would lead away from the claimed invention. *W.L. Gore & Associates, Inc. v. Garlock, Inc.*, 721 F.2d 1540, 220 USPQ 303 (Fed. Cir. 1983), *cert. denied*, 469 U.S. 851 (1984). This information on page 192 of Foster leads away from the claimed subject matter because it says that a major portion of the correct translation result can be provided by the machine as compared with the human translator, in more than a 2:1 ratio (70%/30%). The overall teaching in Foster is that its human translator does not act alone, is always assisted by a translation machine and the machine finishes a translation acceptably

perhaps 70% of the time. Therefore, Appellants submit that Foster cannot reasonably be interpreted to teach translation actually made by a human translator.

Schulz:

Schulz does not and cannot cure this deficiency in Foster because Schulz does not even teach translation, only transcription - from audio language A to textual language A, and only for correcting errors in its prior machine-generated translation from a human source language into resultant language A. Therefore, Foster operation in *arguendo* combination with the audio disclosure and error-correcting transcription disclosure in Schulz does not read on "receiving translation actually made by the user of the portion of the audio signal" as recited in Appellants' claim 1 at least because such combination still does not describe a translation actually made by the user [human translator]. As noted above, the translation is actually made in Foster by the machine-human combination, with the machine typically providing more than twice as much translation (70%) than that provided by the human (30%).

Thus, this limitation of claim 1 and, therefore, claim 1 itself, is not disclosed or suggested by Foster and Schulz taken individually or in any reasonable combination. The 35 U.S.C. 103(a) rejection of claim 1 should be REVERSED and the claim allowed.

Furthermore, with reference to the entire limitation: "receiving translation actually made by the user of the portion of the audio signal," Appellants direct attention to the recited "portion of the audio signal" (emphasis added) part of the limitation. A vocal utterance which is a full

word is capable of translation because a full word has meaning. But an utterance which is less than a full word has ambiguous meaning at best, or no meaning at all.

For example, consider any word, such as, "patent." If an audio/vocal utterance presented only the sound equivalent of "pa" or "pat" to the ears of a human translator, less than the full word, Appellants submit that there cannot be a reliably-accurate word-translation made from only that input. Any other word that starts with vocal sounds represented by those letters could be chosen. For example, if context is provided,⁴ the translated word based on only that utterance could be patented, patent application, patentable, patentably, patenting; etc. If there is no context given, the translated word could be any of the foregoing as well as, e.g., patch, patella, paternal, patio, patrol, patsy, patter, pattern, patty-cake etc. Therefore, it should be apparent that in the recited limitation "receiving translation actually made by the user of the portion of the audio signal" the word "portion" refers to a sound that is properly translatable in the first place. That sound is a full word, at a minimum.

Therefore, because (1) a full audio word is needed as a minimum to read on "portion" in the claim limitation and (2) Foster does not teach human translation of a full word but, rather, teaches human translation of only a part of each word on a keystroke basis until the machine finishes the translation of that word, the combination of Foster and Schulz again fails to disclose or suggest at least "receiving translation actually made by the user of the portion of the audio signal" (emphasis added) as recited in claim 1. For this additional reason, the 35 U.S.C § 103(a)

⁴ In the Office Action, page 4, the Examiner suggests that a human translator can take into account context, grammar, and semantic sense but clearly there remains ambiguity as shown by the various patent-related words that could possibly fill the bill.

rejection of claim 1 based on a combination of Foster in view of Schulz should be REVERSED and the claim allowed.

II. CLAIM 1 IS ALLOWABLE BECAUSE FOSTER AND SCHULZ ARE NOT COMBINABLE

The Office Action concedes that Foster does not disclose or suggest subject matter related to the audio signal recited in claim 1. (Office Action, page 7) Appellants agree.

The Office Action then presents Schulz which discloses audio transcription but which has absolutely nothing to do with translation and immediately concludes that, because Schulz (1) mentions in its background section (col. 1, lines 27-34) that automatic speech recognition systems convert spoken language to written text and (2) discloses (col. 5, lines 30-33) the synchronizing of text with a specific spoken word during playback of an audio file, it would be obvious to one of ordinary skill in the art at the time of the invention to combine Schulz with Foster to read on Appellants' subject matter as recited in claim 1. The alleged rationale given is: "it would have been obvious to one of ordinary skill in the art at the time of the invention to use known methods to retrieve a textual representation of an audio signal for translation in Foster, since it would provide automatic transcription, saving transcription costs, (Schulz, column 1 lines 27-34) while enabling a user to provide fast and accurate translation of speech data." (Office Action, pg 7) Appellants respectfully disagree that this is satisfactory rationale at least for the reason that this is no more than a conclusory statement that merely recites advantages offered by Appellants' claimed subject matter, those advantages being apparent in hindsight after

one reads Appellants' claims.

The Office Action then alleges that it would also have been obvious to "combine the known elements of audio and text synchronization with Foster, since the combination would produce the predictable result of enabling the user to quickly and easily translate and edit text displayed on the monitor including identifying and correcting errors, without interruption during playback of the speech from an audio recording, as indicated in Schulz (column 5 lines 55-58)." (Office Action, pgs 7-8) Appellants again respectfully disagree that this is satisfactory rationale for finding obviousness at least for the reason that this is also no more than a conclusory statement that is also merely reciting advantages offered by Appellants' claimed subject matter, those advantages being apparent in hindsight after reading Appellants' claims.

Appellants rely on the recently decided case KSR International Co. v. Teleflex Inc., 550 U.S. ____ (April 30, 2007) (citing In re Kahn, 441 F.3d 977, 988 (Fed. Cir. 2006)), (hereinafter "KSR") where it was held that rejections on obviousness grounds cannot be sustained by mere conclusory statements; instead, there must be some articulated reasoning with some rational underpinning to support the legal conclusion of obviousness. Appellants submit that the above-noted statements in the Office Action do not represent articulated reasoning. The Examiner's purported motivation to combine the cited references is merely conclusory and based on impermissible hindsight. If it were as obvious to have combined the teachings of Foster and Schulz to achieve the alleged "predictable result" as the Office Action represents, Appellants query, as a threshold matter, why that combination was not previously made. After all, the Examiner has conducted a thorough search and, by not finding a description of that combination

within a single reference, has shown that the alleged “predictable result” has apparently not yet been produced in tangible form.

In this connection, MPEP 2141 (III) offers guidance with respect to various rationales to support rejections under KSR. One exemplary rationale is “obvious to try - choosing from a finite number of identified, predictable solutions, with a reasonable expectation of success.” Appellants submit that it is not obvious to try to combine Foster and Schulz for several reasons. First of all, Foster is a machine language-translation system for operating exclusively on text, involving a human operator only for translating the beginning of each source-language word and more if necessary; this reference does not even hint at audio data input. Quite differently, Schulz is a transcribing system for editing exclusively a transcription of audio (voice) with synchronization between the spoken language and the transcription; this reference does not even hint at language-translation or textual data input. Appellants submit that translation between two different languages on the one hand and transcription from one media to another in the same language on the other hand are two very different activities and common sense suggests that there is no motivation to be derived from a reading of either of these references to seek its combination with the other.

In addition, they operate with divergent technologies, where their combination offers no predictable solution and no reasonable expectation of success. There are divergent technologies involved in, and resultant divergent skill requirements needed for handling (1) completion of Foster's partially-translated text via statistical translation and statistical language models into digital signals for further processing, versus (2) conversion of Schulz's audio signals to digital

signals for further processing. Accordingly, one skilled in an audio-signal processing art need not be similarly skilled in a textual-signal processing art, particularly where in-depth knowledge of statistical translation and statistical language models may be needed for machine-assisted human translation. And the reverse is true as well. This clear difference in technologies makes it unlikely, in Appellants' view, that an interested reader of one of these cited references would be motivated as a result of that reading to seek out the other cited reference for combination purposes in order to solve the problem being solved by the subject matter of Appellants' claims.

The initial burden of establishing a *prima facie* basis to deny patentability to a claimed invention always rests upon the Examiner. In re Oetiker, 977 F.2d 1443, 24 U.S.P.Q.2d 1443 (Fed. Cir. 1992). In rejecting a claim under 35 U.S.C. § 103, the Examiner must provide a factual basis to support the conclusion of obviousness. In re Warner, 379 F.2d 1011, 154 U.S.P.Q. 173 (C.C.P.A. 1967). Based upon the objective evidence of record, the Examiner is required to make the factual inquiries mandated by Graham v. John Deere Co., 86 S.Ct. 684, 383 U.S. 1, 148 U.S.P.Q. 459 (1966). KSR International Co. v. Teleflex Inc., 550 U.S. _____ (April 30, 2007). The Examiner is also required to explain how and why one having ordinary skill in the art would have been realistically motivated to modify an applied reference and/or combine applied references to arrive at the claimed invention. Uniroyal, Inc. v. Rudkin-Wiley Corp., 837 F.2d 1044, 5 U.S.P.Q.2d 1434 (Fed. Cir. 1988). In view of the differences between the references that have been presented herein, Appellants respectfully submit that the Examiner has not met these standards; for example, in this instance, the Office Action has not presented sufficient explanation of how and why one having ordinary skill in the art would have been

realistically motivated to modify either applied reference and/or combine these applied references to attempt to arrive at the claimed subject matter. The Office Action merely presents advantages which become appreciated after a reading of Appellants' claims. (Moreover, *arguendo*, even if they were combinable which Appellants refute, Schulz still would not cure the deficiency of Foster.)

It is established law that one "cannot use hindsight reconstruction to pick and choose among isolated disclosures in the prior art to deprecate the claimed invention." *Ecolchem, Inc. v. Southern Cal. Edison Co.*, 227 F.3d 1361, 1371, 56 USPQ2d 1065 (Fed. Cir. 2000) (citing *In re Fine*, 837 F.2d 1071, 1075, 5 USPQ2d 1780, 1783 (Fed. Cir. 1988)). Indeed, "[c]ombining prior art references without evidence of such a suggestion, teaching, or motivation simply takes the inventor's disclosure as a blueprint for piecing together the prior art to defeat patentability ... the essence of hindsight." *In re Dembiczak*, 175 F.3d 994, 999, 50 USPQ2d 1614, 1617 (Fed. Cir. 1999). Appellants submit that in this instance Appellants' claim 47 was used as such a blueprint to piece together Foster and Schulz. For these reasons, the 35 U.S.C. 103(a) rejection of claim 47 should be REVERSED and the claim allowed.

III. CLAIM 47 IS ALLOWABLE BASED ON ARGUMENTS #I AND #II ABOVE INCORPORATED BY REFERENCE HEREIN BECAUSE SHIOTANI DOES NOT CURE DEFICIENCY OF FOSTER AND BECAUSE SHIOTANI AND SCHULZ ARE NOT COMBINABLE WITH EACH OTHER.

Claim 47, which is rejected on the basis of Foster and Schulz in combination with Shiotani, is also allowable. These references taken individually or in any reasonable combination do not disclose or suggest: "said user actually translating said audio playback of

said information thereby obtaining a translation in a second language, said user using a different section of said GUI to display said translation while making said translation” as recited in claim 47 because of all of the reasons given above for allowability of claim 1 over Foster and Schulz, those reasons being incorporated herein by reference and because Shiotani does not cure any of the deficiencies of Foster and Schulz.

Moreover, Shiotani and Schulz are not properly combinable with each other in the first place. The Office Action, pg 26, notes that Shiotani discloses in Figs. 4(a) and 4(b) a machine translation system where the source string and target string appear side-by-side in the same window. The Examiner then immediately concludes (Office Action, pg 26) that Shiotani, a reference limited to machine translation of text (no audio) is combinable with Schulz, a reference limited to machine transcription of audio (no text). Appellants respectfully disagree.

Appellants rely on the recently decided case KSR International Co. v. Teleflex Inc., 550 U.S. _____ (April 30, 2007) (citing In re Kahn, 441 F.3d 977, 988 (Fed. Cir. 2006)), (hereinafter “KSR”) where it was held that rejections on obviousness grounds cannot be sustained by mere conclusory statements; instead, there must be some articulated reasoning with some rational underpinning to support the legal conclusion of obviousness. Appellants submit that the above-noted statement in the Office Action does not represent articulated reasoning. The Examiner's purported motivation to combine the cited references is merely conclusory and based on impermissible hindsight. If it were as obvious to have combined the teachings of Shiotani and Schulz to achieve the alleged “predictable result” as the Office Action represents, Appellants query, as a threshold matter, why that combination has not previously been made. The answer is

that the combination is actually not obvious, at least because there are multiple differences between the two references including un-related technological disciplines, namely, optical character recognition versus audio technology, and that only after reading Appellants' claims may the combination arguably appear to be obvious. After all, the Examiner has conducted a thorough search and, by not finding a description of that combination within a single reference, has shown that the alleged "predictable result" has apparently not yet been produced in tangible form.

In this connection, MPEP 2141 (III) offers guidance with respect to various rationales to support rejections under KSR. One exemplary rationale is "obvious to try - choosing from a finite number of identified, predictable solutions, with a reasonable expectation of success." Appellants submit that it is not obvious to try to combine Shiotani and Schulz for several reasons. First of all, Shiotani is a machine language-translation system for operating exclusively on text, involving a human operator only for correction purposes; this reference does not even hint at audio data input. Quite differently, Schulz is a transcribing system for editing exclusively a transcription of audio (voice) with synchronization between the spoken language and the transcription; this reference does not even hint at language-translation or textual data input. Appellants submit that translation between two different languages on the one hand and transcription from one media to another in the same language on the other hand are two very different activities and common sense suggests that there is no motivation to be derived from a reading of either of these references to seek its combination with the other.

Indeed, they operate with divergent technologies, where their hypothetical combination

offers no predictable solution and no reasonable expectation of success. The divergent technologies and inherent divergent skill requirements are needed for handling, (1) conversion of Shiotani's textual input via optics to digital signals for further processing, versus (2) conversion of Schulz's audio signals to digital signals for further processing.

For example, Shiotani (col. 2, lines 13-14) discusses an optical character reader (OCR) involving principles based on the physics of optics. Momentarily expanding on this subject for illustrative purposes, OCR is mechanical or electronic translation of images of text into machine-editable text, using optical techniques such as mirrors and lenses in combination with scanners and digital processing. OCR is a process by which glyph images (the visual image of a character) yield character codes. Given a picture of letters arranged as words, OCR is supposed to give back strings of character codes arranged as words. Individual dots of the digital image are represented by a number that varies as function of black through gray to white (for black/white images). Locations of the scan are identified as pixels (picture elements). This brief snippet of OCR information may provide an inkling of what someone with skill in this art has mastered.

By contrast, Schulz (col. 4, lines 50-53) discusses a mu-law encoded eight-bit digital signal. The mu-law algorithm is a companding algorithm, whose purpose is to reduce the dynamic range of an audio signal. In the analog domain, this can increase the signal-to-noise ratio achieved during transmission, and in the digital domain it can reduce quantization error. Beyond this, speech recognition involves many considerations such as complexity of the language model. By this is meant the number of permissible words following each word. The

simplest language model can be specified as a finite-state network. One measure of the difficulty of the task of combining vocabulary size and language model is called "perplexity" which is the geometric mean of the number of words that can follow a word, after a language model has been applied. This does not begin to scratch the surface of the subject of speech recognition, but this brief snippet of speech recognition information may provide an inkling of what someone with skill in this art has mastered.

Appellants have juxtaposed the above two paragraphs to clearly show that the subjects discussed therein are mutually exclusive. One topic has virtually nothing to do with the other. Accordingly, one skilled in the audio signal processing art need not be similarly skilled in the text signal processing and vice-versa. This clear difference in audio/textual technologies, in addition to the translation vs. transcription difference noted above, make it unlikely, in Appellants' view, for a reader of either one of these references to find any motivation within it to combine it with the other.

The initial burden of establishing a *prima facie* basis to deny patentability to a claimed invention always rests upon the Examiner. In re Oetiker, 977 F.2d 1443, 24 U.S.P.Q.2d 1443 (Fed. Cir. 1992). In rejecting a claim under 35 U.S.C. § 103, the Examiner must provide a factual basis to support the conclusion of obviousness. In re Warner, 379 F.2d 1011, 154 U.S.P.Q. 173 (C.C.P.A. 1967). Based upon the objective evidence of record, the Examiner is required to make the factual inquiries mandated by Graham v. John Deere Co., 86 S.Ct. 684, 383 U.S. 1, 148 U.S.P.Q. 459 (1966). KSR International Co. v. Teleflex Inc., 550 U.S. _____ (April 30, 2007). The Examiner is also required to explain how and why one having ordinary skill in

the art would have been realistically motivated to modify an applied reference and/or combine applied references to arrive at the claimed invention. Uniroyal, Inc. v. Rudkin-Wiley Corp., 837 F.2d 1044, 5 U.S.P.Q.2d 1434 (Fed. Cir. 1988). In view of the differences between the references that have been presented herein, Appellants respectfully submit that the Examiner has not met these standards; for example, in this instance, the Office Action has not presented sufficient explanation of how and why one having ordinary skill in the art would have been realistically motivated to modify either applied reference and/or combine these applied references to arrive at the claimed subject matter. The Office Action merely presents advantages which become appreciated after a reading of Appellants' claims.

It is established law that one "cannot use hindsight reconstruction to pick and choose among isolated disclosures in the prior art to deprecate the claimed invention." *Ecologchem, Inc. v. Southern Cal. Edison Co.*, 227 F.3d 1361, 1371, 56 USPQ2d 1065 (Fed. Cir. 2000) (citing *In re Fine*, 837 F.2d 1071, 1075, 5 USPQ2d 1780, 1783 (Fed. Cir. 1988)). Indeed, "[c]ombining prior art references without evidence of such a suggestion, teaching, or motivation simply takes the inventor's disclosure as a blueprint for piecing together the prior art to defeat patentability – the essence of hindsight." *In re Dembiczak*, 175 F.3d 994, 999, 50 USPQ2d 1614, 1617 (Fed. Cir. 1999). Appellants submit that in this instance Appellants' claim 47 was used as such a blueprint to piece together Shiotani and Schulz.

Shiotani was cited against claim 47 to cure the Schulz deficiency of not displaying a textual representation in a split screen in a translation window. That deficiency is not cured because the references cannot be combined for the reasons given above. For this additional

reason the 35 U.S.C. § 103(a) rejection of claim 47 should be REVERSED and the claim allowed. (Moreover, *arguendo*, even if they were combinable which Appellants refute, Shiotani still would not cure the deficiency of Foster.)

IV: INDEPENDENT CLAIMS 20, 21, 40 AND DEPENDENT CLAIMS:

Each one of independent claims 20, 21 and 40, contains a limitation which is the same as, or similar to, that limitation of claim 1 upon which argument for allowability of claim 1 was focused. Therefore, claims 20, 21 and 40 are likewise allowable and stand or fall with claim 1.

Dependent claims 2-19, dependent from claim 1, are allowable, at least for reasons based on their respective dependencies from allowable base claim 1.

Dependent claims 22-39, dependent from claim 21, are allowable, at least for reasons based on their respective dependencies from allowable base claim 21.

Dependent claims 41-46, dependent from claim 40, are allowable, at least for reasons based on their respective dependencies from allowable base claim 40.⁵

⁵ Claims 41 and 46 were rejected on the basis of Foster, Schulz and Saindon, the last reference of which has not been previously addressed in this brief. Suffice it to say that Saindon was cited only against claims 41 and 46, and merely to allegedly disclose a system for automated transcription and translation that processes text to visually distinguish the names of people, places and organization using a word processor. "The system processes the text to determine if all proper nouns are capitalized...." (Office Action, pg 21) Saindon does not cure any deficiencies noted herein with respect to Foster, Schulz or Shiotani.

CONCLUSION

Appellants respectfully request that the Honorable Board REVERSE the final rejection of the appealed claims.

To the extent necessary, a petition for an extension of time under 37 C.F.R. § 1.136 is hereby made. Please charge any shortage in fees due in connection with the filing of this paper, including extension of time fees, to Deposit Account No. 07-2347 and please credit any excess fees to such deposit account.

Respectfully submitted,

/Joel Wall/
Joel Wall - Registration 25,648

Date: September 16, 2009

c/o Eddy A. Valverde, Patent Paralegal
Verizon Patent Management Group
1320 North Courthouse Road, 9th Floor
Arlington, VA 22201 - 2909
Tel: 703.351.3032
Fax: 703.351.3665
Customer No. 25,537

VIII: CLAIMS APPENDIX

1. A method for facilitating translation of an audio signal that includes speech to another language, comprising:
 - retrieving a textual representation of the audio signal;
 - presenting the textual representation to a user;
 - receiving selection of a segment of the textual representation for translation;
 - obtaining a portion of the audio signal corresponding to the segment of the textual representation;
 - providing the segment of the textual representation and the portion of the audio signal to the user; and
 - receiving translation actually made by the user of the portion of the audio signal.
2. The method of claim 1, wherein the retrieving a textual representation includes:
 - generating a request for information,
 - sending the request to a server, and
 - obtaining, from the server, at least the textual representation of the audio signal.
3. The method of claim 1, wherein the presenting the textual representation to a user, includes:

obtaining the audio signal,
providing the audio signal and the textual representation of the audio signal to the user,
and
visually synchronizing the providing of the audio signal with the textual representation of the audio signal.

4. The method of claim 3, wherein the obtaining the audio signal includes:
accessing a database of original media to retrieve the audio signal.

5. The method of claim 3, wherein the obtaining the audio signal includes:
receiving input, from the user, regarding a desire for the audio signal,
initiating a media player, and
using the media player to obtain the audio signal.

6. The method of claim 1, wherein the receiving selection of a segment of the textual representation includes:
identifying a portion of the textual representation selected by the user,
accessing a server to obtain text corresponding to the portion of the textual representation, and
receiving, from the server, the text corresponding to the portion of the textual representation.

7. The method of claim 6, wherein the text includes a transcription of the audio signal and metadata corresponding to the portion of the textual representation.
8. The method of claim 1, wherein the obtaining a portion of the audio signal includes:
- initiating a media player, and
 - using the media player to obtain the portion of the audio signal.
9. The method of claim 8, wherein the using the media player includes:
- identifying, by the media player, the segment of the textual representation, and
 - retrieving the portion of the audio signal corresponding to the segment of the textual representation.
10. The method of claim 9, wherein the identifying the segment includes:
- identifying time codes associated with a beginning and an ending of the segment of the textual representation.
11. The method of claim 9, wherein the segment of the textual representation includes a starting position in the textual representation; and
- wherein the identifying the segment includes:

identifying a time code associated with the starting position in the textual representation.

12. The method of claim 1, wherein the providing the segment of the textual representation and the portion of the audio signal to the user includes:

displaying the segment of the textual representation in a same window as will be used by the user to provide the translation of the portion of the audio signal.

13. The method of claim 1, wherein the providing the segment of the textual representation and the portion of the audio signal to the user includes:

visually synchronizing the providing of the portion of the audio signal with the segment of the textual representation.

14. The method of claim 13, wherein the segment of the textual representation includes time codes corresponding to when words in the textual representation were spoken.

15. The method of claim 14, wherein the visually synchronizing the providing of the portion of the audio signal with the segment of the textual representation includes:

comparing times corresponding to the providing of the portion of the audio signal to the time codes from the segment of the textual representation, and

visually distinguishing words in the segment of the textual representation when the words are spoken during the providing of the portion of the audio signal.

16. The method of claim 1, wherein the providing the segment of the textual representation and the portion of the audio signal to the user includes:

permitting the user to control the providing of the portion of the audio signal.

17. The method of claim 16, wherein the permitting the user to control the providing of the portion of the audio signal includes:

allowing the user to at least one of fast forward, speed up, slow down, and back up the providing of the portion of the audio signal using foot pedals.

18. The method of claim 16, wherein the permitting the user to control the providing of the portion of the audio signal includes:

permitting the user to rewind the portion of the audio signal at least one of a predetermined amount of time and a predetermined number of words.

19. The method of claim 1, further comprising:

publishing the translation to a user-determined location.

20. A system for facilitating translation of speech between languages, comprising:

means for obtaining a textual representation of the speech in a first language;

means for presenting the textual representation to a user;

means for receiving selection of a portion of the textual representation for translation;
means for retrieving an audio signal in the first language that corresponds to the portion of the textual representation;
means for providing the portion of the textual representation and the audio signal to the user; and
means for receiving translation actually made by the user of the audio signal into a second language.

21. A translation system, comprising:

a memory configured to store instructions; and

a processor configured to execute the instructions in memory to:

obtain a transcription of an audio signal that includes speech,

present the transcription to a user,

receive selection of a portion of the transcription for translation,

retrieve a portion of the audio signal corresponding to the portion of the transcription,

provide the portion of the transcription and the portion of the audio signal to the user, and

receive from the user a translation actually made by the user of the portion of the audio signal.

22. The system of claim 21, wherein when obtaining a transcription, the processor is configured to:

- generate a request for information,
- send the request to a server, and
- obtain, from the server, at least the transcription of the audio signal.

23. The system of claim 21, wherein when presenting the transcription to a user, the processor is configured to:

- obtain the audio signal,
- provide the audio signal and the transcription of the audio signal to the user, and
- visually synchronize the providing of the audio signal with the transcription of the audio signal.

24. The system of claim 23, wherein when obtaining the audio signal, the processor is configured to:

- access a database of original media to retrieve the audio signal.

25. The system of claim 23, wherein when obtaining the audio signal, the processor is configured to:

- receive input, from the user, regarding a desire for the audio signal,
- initiate a media player, and

use the media player to obtain the audio signal.

26. The system of claim 21, wherein when receiving selection of a portion of the transcription, the processor is configured to:

identify a range of the transcription selected by the user,
access a server to obtain text corresponding to the range of the transcription, and
receive, from the server, the text corresponding to the range of the transcription.

27. The system of claim 26, wherein the text includes metadata corresponding to the range of the transcription.

28. The system of claim 21, wherein when retrieving a portion of the audio signal, the processor is configured to:

initiate a media player, and
use the media player to obtain the portion of the audio signal.

29. The system of claim 28, wherein the media player is configured to:

identify the portion of the transcription, and
retrieve the portion of the audio signal corresponding to the portion of the transcription.

30. The system of claim 29, wherein when identifying the portion, the media player is

configured to:

identify time codes associated with a beginning and an ending of the portion of the transcription.

31. The system of claim 29, wherein the portion of the transcription includes a starting position in the transcription; and

wherein when identifying the portion, the media player is configured to:

identify a time code associated with the starting position in the transcription.

32. The system of claim 21, wherein when providing the portion of the transcription and the portion of the audio signal to the user, the processor is configured to:

present a split screen in a translation window, the translation window including a translation section and a transcription section, and

display the portion of the transcription in the transcription section.

33. The system of claim 21, wherein when providing the portion of the transcription and the portion of the audio signal to the user, the processor is configured to:

visually synchronize the providing of the portion of the audio signal with the portion of the transcription.

34. The system of claim 33, wherein the portion of the transcription includes time

codes corresponding to when words in the transcription were spoken.

35. The system of claim 34, wherein when visually synchronizing the providing of the portion of the audio signal with the portion of the transcription, the processor is configured to:

compare times corresponding to the providing of the portion of the audio signal to the time codes from the portion of the transcription, and

visually distinguish words in the portion of the transcription when the words are spoken during the providing of the portion of the audio signal.

36. The system of claim 21, wherein when providing the portion of the transcription and the portion of the audio signal to the user, the processor is configured to:

permit the user to control the providing of the portion of the audio signal.

37. The system of claim 36, further comprising:

foot pedals configured to aid the user to at least one of fast forward, speed up, slow down, and back up the providing of the portion of the audio signal.

38. The system of claim 36, wherein when permitting the user to control the providing of the portion of the audio signal, the processor is configured to:

permit the user to rewind the portion of the audio signal at least one of a predetermined

amount of time and a predetermined number of words.

39. The system of claim 21, wherein the processor is further configured to:
publish the translation to a user-determined location.

40. A graphical user interface, comprising:
a transcription section that includes a transcription of non-text information in a first language;
a translation section that receives a translation actually made by the user of the non-text information into a second language; and
a play button that, when selected, causes:
retrieval of the non-text information to be initiated,
playing of the non-text information, and
the playing of the non-text information to be visually synchronized with the transcription in the transcription section.

41. The graphical user interface of claim 40, wherein the transcription visually distinguishes names of people, places, and organizations.

42. The graphical user interface of claim 40, further comprising:
a configuration button, that when selected, causes a window to be presented, the window

permitting an amount of backup to be specified, the amount of backup including one of a predetermined amount of time and a predetermined number of words.

43. The graphical user interface of claim 42, wherein the window further permits a name to be given for the translation and a location of publication to be specified.

44. The graphical user interface of claim 40, wherein the play button further causes words in the transcription to be visually distinguished in synchronism with the words in the non-text information being played.

45. The graphical user interface of claim 40, wherein the non-text information includes at least one of audio and video.

46. The graphical user interface of claim 40, wherein the graphical user interface is associated with a word processing application.

47. A method, comprising:
a user listening to an audio playback of information in a first language while viewing a textual transcription of said information in said first language on a transcription section of a graphical user interface (GUI), said textual transcription being synchronized with said audio playback; and

said user actually translating said audio playback of said information thereby obtaining a translation in a second language, said user using a different section of said GUI to display said translation while making said translation,

whereby the synchronizing of said audio playback with said textual transcription aids said user in making said translation.

IX. EVIDENCE APPENDIX

None.

X. RELATED PROCEEDINGS APPENDIX

None.